

COMP 4432

Machine Learning

Group Project

Group 20

Student: Leung Kit Chuen

Student ID: 20051248D

Personal Contribution

o What is your main role in doing this project?

In this project, I am mainly responsible for several tasks: model debugging, designing random sampling function, improving the model performance using different techniques, analyzing the model performance, and documenting.

As for designing the random sampling function, it is used for rebalancing the dataset to reduce the bias within the model. By doing so, our group successfully makes the model performance fair without an accuracy paradox.

Regarding the model improvement, I mainly focus on designing the grid search and bagging method (ensemble model).

o What other ideas you have considered to formulate the problem? Have your team members accepted or rejected them?

Initially, our group selected the logistic regression model as the target to improve its performance further. However, I have found that the model could not improve after adopting many technical methods. To cope with this problem, I suggest our group find an alternative model to implement the improvement process. Finally, we came up with using the kNN model as the second targeted model for improvement.

o What were the main difficulties encountered by the whole team? and by you individually?

The main difficulty our group encountered was the computational time of model training. To cope with this challenge, we have leveraged the computer's GPU in the Industrial Center to boost our model training progress.

o What have you learned from this project?

In this project, I have learned that it is essential to improve the dataset and model architecture to obtain better performance improvement.